

## Science Gateways: The Long Road to the Birth of an Institute

Sandra Gesing  
University of Notre Dame  
Notre Dame, USA  
[sandra.gesing@nd.edu](mailto:sandra.gesing@nd.edu)

Nancy Wilkins-Diehr  
University of California  
San Diego, USA  
[wilkinsn@sdsc.edu](mailto:wilkinsn@sdsc.edu)

Maytal Dahan  
University of Texas  
Austin, USA  
[maytal@tacc.utexas.edu](mailto:maytal@tacc.utexas.edu)

Katherine Lawrence  
University of Michigan  
Ann Harbor, USA  
[kathla@umich.edu](mailto:kathla@umich.edu)

Michael Zentner  
Purdue University  
West Lafayette, USA  
[mzentner@purdue.edu](mailto:mzentner@purdue.edu)

Marlon E. Pierce  
Indiana University  
Bloomington, USA  
[marpierce@iu.edu](mailto:marpierce@iu.edu)

Linda B. Hayden  
Elizabeth City State University  
Elizabeth City, USA  
[haydenl@mindspring.com](mailto:haydenl@mindspring.com)

Suresh Marru  
Indiana University  
Bloomington, USA  
[smarru@iu.edu](mailto:smarru@iu.edu)

### Abstract

*Nowadays, research in various disciplines is enhanced via computational methods, cutting-edge technologies and diverse resources including computational infrastructures and instruments. Such infrastructures are often complex and researchers need means to conduct their research in an efficient way without getting distracted with information technology nuances. Science gateways address such demands and offer user interfaces tailored to a specific community. Creators of science gateways face a breadth of topics and manifold challenges, which necessitate close collaboration with the domain specialists but also calling in experts for diverse aspects of a science gateway such as project management, licensing, team composition, sustainability, HPC, visualization, and usability specialists. The Science Gateway Community Institute tackles the challenges around science gateways to support domain specialists and developers via connecting them to diverse experts, offering consultancy as well as providing a software collaborative, which contains ready-to-use science gateway frameworks and science gateway components.*

### 1. Introduction

Billions of people use the web every day. From banking to travel arrangements to connecting with family, the web has changed how we conduct our lives. Its impact on scientific research is no different. Science gateways are used by many researchers to conduct their work. Implemented as web, desktop, and mobile-device applications, science gateways provide

community-based access to shared, distributed resources and services - telescopes, sensor arrays, supercomputers, digital repositories, software as a service, collaboration environments, and more. Gateways enable the formation of scientific communities, accelerating and transforming the discovery process, and engaging citizens and students in the scientific process. They represent a fundamental social and technological change in how science is being conducted.

Since gateways form end-to-end solutions tailored to the communities' needs, the creators of science gateways are concerned with a diversity of topics and have to tackle various challenges. The fundamental first step is to understand the requirements of the community and an at least high-level insight into the research area. The following domain-related topics need to be addressed in close collaboration with a target community.

- Specific goal of a science gateway
- Visions/demands on the layout
- Priorities of features and options, e.g., a list from must-have to great-to-have options
- Integration of existing applications or development of applications
- Technologies of the applications
- Demand on computational resources for an efficient and effective use of an application
- Visualization of results and job submissions
- Security demands, i.e., users may want to share data after they received a patent or published it
- Need for workflows

- Data management, e.g., the location of input data, the demand on storage for input, intermediate and output data

When these topics are clarified, creators should consider further topics for a seamless integration of cyberinfrastructure (CI) to achieve efficiently a sustainable solution.

- Available infrastructure including security infrastructure and resources
- Available support of suitable technologies
- Scalability of suitable technologies
- Effort for extending existing technologies compared to novel developments
- Synergy effects with other science gateway projects
- Experience with available solutions and/or programming languages

National Science Foundation (NSF) reports [1,2,3,4,5,6,7] describe the enormous effect that CI is having on science, but they do not highlight the dramatic shift taking place in how the average user accesses advanced CI. Specialized, technical interfaces such as the Unix command line no longer serve the broad, diversifying community. Because research challenges are more interdisciplinary, scientists demand the ability to focus on their research questions, as well as to convey their results to the general public. Moreover, faculty wants to more efficiently and effectively integrate their research with teaching. Programming skills amongst interdisciplinary teams can vary widely. For all of these reasons, science teams are turning to the web [7,8]. But the significance of gateways goes beyond replacement of command-line tools to include sharing of research (methods, data, visualizations, etc.), increasing transparency of methods and data used, reproducing others' results, shielding users from much of the complex underlying infrastructure, and applying research-grade tools in classrooms and training.

## 2. Background

Over the past ten years, research, workshops, collected papers, and special journal issues [9, 10] have called attention to the challenges and successes of science gateways, while extending our understanding of how to address these challenges. Despite notable successes, building and operating gateways is still too costly and too prone to failure. The success rate and capabilities of gateways must be further increased while reducing associated costs and effort spent. The next step is to leverage and enhance every dollar that is invested in gateways by supporting best practices for development, deployment, and sustainability.

Gateway development has often been done in an *ad-hoc* way, limiting success and long-term impact. The science gateways name originated from a focus area within the NSF TeraGrid program (2005-2011) [11, 12]. This program developed policies and procedures that made the use of high performance computers via the web viable. Gateway developers in the program worked together both to design the policies and procedures and then use them as they incorporated the use of supercomputers in their own gateways. But the program had surprising, unintended benefits. In addition to working on the common goal to use supercomputers in gateways, developers benefitted from the mere existence of a community designing advanced web portals for science. There was unexpected commonality across disciplines and few other forums for sharing experiences in scientific development.

In the course of this work, we noticed problems with duplication of effort and the long-term sustainability of gateways. Science gateways enable research, sometimes for thousands, but are not research projects in and of themselves. As a result, they struggle to find the right avenues for sustainable funding [13]. This work revealed a key finding: Gateways funded via short-term research grants can follow a *destructive* cycle. Prototypes are developed, early adopters are identified, and interest in using the resulting gateway is encouraged. Then the project ends. Understandably, researchers who had invested time in gateway use become disillusioned and less likely to use gateways again.

A 2008 white paper [13] highlighted these challenges and led to an NSF-funded study (2009-2011) to investigate what contributes to gateway success [14]. Participants in focus groups helped identify solutions. One outcome is that gateway builders need a common, reliable reservoir of software, expertise, and support they can call on to increase their chances of successful development and implementation. Developers should be able to easily share experience and use community-contributed tools to more quickly create robust, less expensive, and sustainable gateways. Their efforts should maximize the likelihood that their respective communities become interested in and maintain those gateways when initial funding ends. Similarly, gateway software and service providers need to plan for the future with clear leadership to more effectively serve the gateway community.

### 3. Large-Scale Survey

A further award (2012-2016) allowed to explore via a 5000-respondent survey [15] whether the ideas gleaned from small but carefully chosen focus groups resonated with the larger research community. To our best knowledge the large-scale survey is the most comprehensive of its kind, covering the breadth of the scientific and engineering disciplines represented at NSF and across all geographic areas in the U.S. The survey targeted NSF principal investigators, gateway developers, and leaders in higher education (e.g., CIOs, CTOs). The high degree of community interest in gateways was reflected by the unexpectedly large number of responses. Out of 29,000 invitations, we received nearly 5,000 responses, a 17% response rate, considered quite high for these types of surveys [16,17,18]. Some 88% of researchers and educators indicated that science gateways were important to their work.

The survey contains several key inquiries such as the importance of gateways across domains. Who was using gateways and for what? Who was developing gateways and for what? What types of skills did they have on their projects? What did they wish they had? What types of training is preferred by developers? What influences decisions to use a piece of software? Finally, the survey asked which kind of broader services would be interesting. Survey questions and anonymized responses are available at <http://hdl.handle.net/2027.42/110982>.

Portions of the survey results have been featured in prior publications. Details on the survey design, population sample and implementation as well as some early findings were featured at the 2014 Gateway Computing Environments workshop [15]. In that paper we describe the roles survey respondents play in gateway development (most were PIs given the population sample). We also asked about the types of people PIs had or wished they had on their gateway projects. 65% of the projects involved students. Respondents were closely split amongst staff they wish they had - quality assurance (39%), graphic designer (34%), usability consultant (32%) and professional software developer (29%).

Of 23 proposed Institute services, 20 received more than a 50% interest rating. So, first we gathered ideas from key focus group participants, then we vetted these with the much larger research community. Finally, we implemented these ideas as services for the community via the Science Gateways Community Institute. An important theme in the responses was the cross-cutting capabilities that gateways bring to the spectrum of NSF-funded activities.

### 4. Institute Components

The Science Gateways Community Institute is the result of years of study and community input. An organization at this level can serve as a focal point that galvanizes the community, leverages the extensive but disjointed investments in gateway resources, codes, and expertise, and supports longer-term career paths for gateway developers by highlighting the importance of gateways in the conduct of research today. The Institute's mission is to provide resources, expertise, community support, and education to speed development and application of more robust, less expensive, and more sustainable gateways that serve science and engineering research and education. By doing so, the Institute will enable diverse scholarly communities to focus on their research and educational objectives.

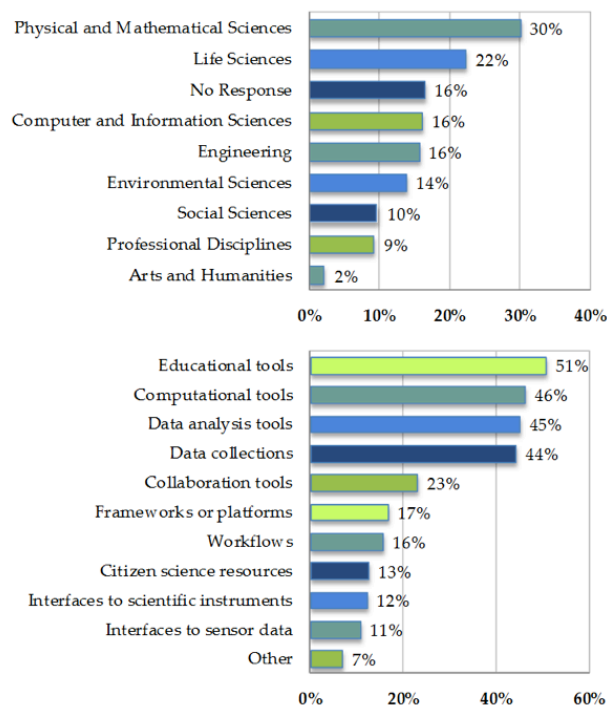


Figure 1: Survey respondent domains (top) and areas of software development (bottom)

The Institute is composed of five complementary areas, as depicted in Figure 2 and described in more detail in subsequent sections. Collectively, these five Institute areas support a community of practice that consists of people who share an interest in a topic and expand their expertise and knowledge through ongoing interactions [16]. These communities are capable of adapting to the ever-evolving tacit and explicit knowledge necessary for their work. As a steward of this community, the Institute will likewise change dynamically through interactions with clients.

Our target science gateway community consists of three constituent groups: (1) developers of gateways and related software, including programmers and technical specialists focused on gateway design, production, or maintenance; (2) domain-based scientists and engineers who need gateways to support their research and teaching; and (3) campus-based administrators or IT staff who provide technology services to their institutions but who may lack the manpower or expertise to deliver specialized gateways.

In the following sections, we describe the Institute's areas as they have been implemented at the start of funding. We anticipate that service offerings will be adapted and expanded according to the communities' feedback.

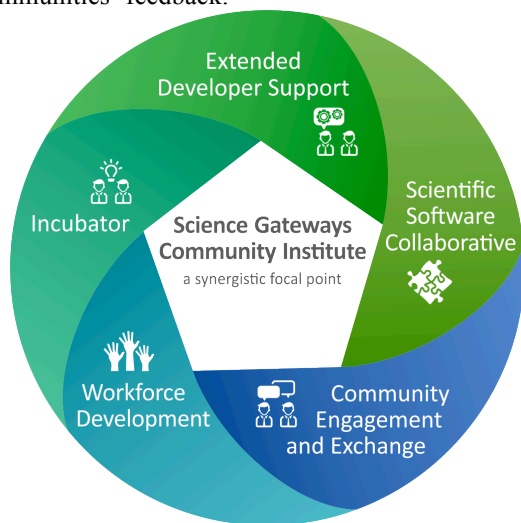


Figure 2: Conceptual overview on the Institute

#### 4.1. Incubator

The *Incubator* will combine aspects of a business incubator and an open-source software incubator. With respect to gateway development, such experience includes not just how to implement the technical details, but all aspects of the gateway lifecycle from planning and design through sustainable operations, particularly:

*Gateway planning and design:* These clients are contemplating gateway development in advance of or in response to a solicitation.

*Gateway project ramp-up:* These clients have funding and are in the process of hiring and organizing staff to execute their projects.

*Active operations:* These are fully staffed clients already executing their plans.

*Transition:* With initial funding nearing its end, these clients need to make decisions about submitting renewal proposals, identifying alternate sources of funding, or ramping down their activities.

At each stage of a client's efforts, we offer a variety of services, each of which was determined valuable to the community by respondents to the survey. One function of the Incubator, as in business incubators, will be the formation of cohorts of people that interact with each other after they undergo initial incubation training [20]. In addition, we expect "graduates" of our program to work with incoming clients as advisors, leading to new research collaborations that might not have formed otherwise.

We offer cohort training sessions plus individualized incubator-client follow-on consultation. The services will be divided into three functional areas: (1) technology planning services, (2) business planning services, and (3) client interaction planning.

*Technology Planning Services:* In the survey, the activity of "choosing/adapting to technologies" was a highly ranked need by administrators and application creators (78% and 66%, respectively). Four key services within the *Incubator* will be (1) choosing tools and methods to support software development, (2) choosing components and technologies to implement a particular client project, (3) security planning, and (4) data management consulting. The goal is to equip our clients with sufficient knowledge about their options as they embed tools and components in their own processes and then, as they begin execution, assist them with individualized consulting services. Helping clients choosing technologies depends on the specifics of each project. Thus, they are encouraged to evaluate and describe the needs of their user communities prior to choosing technologies.

*Business Planning Services:* These services assist clients with strategic decision making and assessing the impact of strategic directions on operational costs. They involve aspects of startup business planning that were supported by our survey results, including project management (supported by 52% of administrators and 38% of application creators), licensing (70% of administrators and 61% of application creators), and staff and sustainability planning (79% of administrators and 62% of application creators). While most clients are not and will not be commercial startup businesses, many of the exercises undertaken by such startups are useful for non-commercial initiatives. Those teams that have commercialization as one of their sustainability goals will find that Incubator services are valuable precursors to the more rigorous and commercially oriented programs such as NSF I-Corps™ as they seek to further such a strategy.

Business planning helps client teams crystallize their messages in the form of a short description that is understandable by those who may not be experts in the client teams' fields, helping our clients consider the

essence of what they want to accomplish and differentiate it from other gateway efforts.

This activity will be followed by completion of a business model canvas [25]. This tool will ask project members to consider topics like value propositions, key partners, key activities, customer relationships, product users, needed resources, cost structure, and anticipated revenue streams. The business model canvas will encourage project stakeholders to focus on concrete challenges and dependencies they must address. Subsequently, our clients will be engaged in goal-oriented planning to move them from traditional, “percentage allocation,” research-project thinking toward task- and milestone-driven thinking, where funding allocations are mapped to goal achievement.

Project-management expertise in the Incubator helps clients in the planning stage construct milestone-based, high-level project plans and resource allocations and advise those beginning operations on how to create execution plans that achieve stated milestones using resources identified in the financial plan.

The licensing component helps principal investigators and management teams to make sensible licensing choices for (1) their resulting gateways, (2) software that may be contributed by others to their gateways, and (3) data or other computational products produced by their gateways.

The internal-support-building component works with clients to develop strategies and best practices to build internal support in their institutions for their efforts, especially at institutions with multiple gateway efforts seeking to create a stable of in-house expertise. This work involves defining, to institutional leadership, the value propositions of synergetic effects between existing projects as justification to leverage additional funding, staffing, or executive support.

The sustainability-planning component builds on the other components. Project leaders are encouraged to shift their mindset from executing the original project plan to supporting a continuous cycle of intellectual property creation and service delivery. Participants in sustainability planning examine the strengths and challenges of their projects, the unique positions they occupy (or could occupy), their ongoing financial needs, and best options for financial support.

Finally the Incubator supports two areas where our survey showed strong interest. We offer usability studies on software and CI (of interest to 65% of administrators and 66% of application creators) and impact-measurement planning (of interest to 72% of administrators and application creators) to help each client design appropriate impact measures and efforts to capture these measures using a combination of qualitative (periodic inter-personal feedback mechanisms) and quantitative methods (web activity,

citations, usage patterns [27,28], trends, and altmetrics [29]).

## 4.2 Extended Developer Support

Building gateways in collaboration with Institute clients is a cornerstone activity of the project. Clients may be identified through the Community Engagement and Exchange component, learn the basics of project and software management through the Incubator component, and use software-as-a-service through the Scientific Software Collaborative. The Extended Developer Support area is where the Institute provides hands-on, individualized, embedded support in collaboration with clients to create new gateways and expand the capabilities of existing ones.

The goals of Extended Developer Support are to help potential gateways come into existence; help existing gateways adapt to new resources, technologies, and user communities; provide “burst” support to help gateways with smaller issues; and develop deep understanding of community needs that feed back into and guide other Institute areas. Extended Developer Support provides in-depth technical and scientific understanding of the science gateway community’s challenges and requirements, leading to, for example, co-authored papers between Institute members and clients. Finally, Extended Developer Support will contribute to the Institute’s broader impact through collaboration with the Workforce Development area to help train the next generation of cyberinfrastructure developers.

Extended Developer Support is inspired by related efforts in the XSEDE science gateway program, which provides developer support to help integrate a gateway with XSEDE. Such staff time is granted through a peer-reviewed allocation process. Staff developers are distributed throughout XSEDE and bring expertise on many approaches to building science gateways. In the European Union, the Scientific Gateway Based User Support (SCI-BUS) [30] activity had a similar aim but based support on a common software framework, WS-PGRADE/gUSE [31].

Our support is different in detail from both of these efforts. Unlike the XSEDE gateway program, we do not simply serve as a gateway integrator to XSEDE resources. Through Extended Developer Support, we work with client gateways on interfaces to a range of resources such as campus clusters, computing clouds, data collections, instruments, and sensor networks. Gateway support will fall into three categories:

*Short-term consultations* (including help desk-like support and engagements requiring up to one month’s effort) supplement efforts in the Incubator and be assigned dynamically. These consultations may be

useful for gateway developers during initial work with the *Incubator* or as ongoing support in concert with the *Incubator*.

*Long-term consultations* will be 2–12 months and involve 25–50% of an Extended Developer Support staff member's effort per gateway. These effort levels are based on current XSEDE extended collaborative support effort guidelines. These consultations will support major science gateway efforts.

*Collaborative consultations* is work funded externally in which the client gateway provides funds to the Institute. The level of effort is determined on a case-by-case basis. Collaborative consultations are an important part of the Institute's own sustainability strategy.

While the Institute's team members have extensive experience with constituent packages (e.g., HUBzero [23][28], Apache Airavata [22][36], CyVerse [35]), support will not be limited to those solutions. Extended Developer Support also helps gateway development teams use open-source solutions such as Galaxy and Jupyter Notebooks and act as an integrator with relevant services such as Globus Transfer and Globus Auth. The Institute also engages with more data-centric gateways, complementing the computing-centric gateways that our team members primarily serve today.

Extended Developer Support efforts will evolve over time. We work closely with Community Engagement and *Exchange* and the *Incubator* to identify candidate clients for extended support, beginning with the more than 1,500 interested contacts from our survey. We are in the design phase of the Institute and are defining a process to evaluate and select from the potentially large number of projects for allocated support. This is complementary to the XSEDE process, provided as a service for those already making XSEDE requests. We outline here our initial process but expect to improve it iteratively with experience. Institute leadership makes support allocation decisions through an internal review that scores the project based on its intellectual merit (potential for publications), broader impact (size and type of community it will serve), transformative potential (such as requirements for novel infrastructure), and management maturity (as evaluated through the *Incubator*). The resulting evaluation is used to determine the priority and level of effort for the consultation.

Extended Developer Support by itself is not designed to be scalable to more than 10–12 clients per year, but by cooperation with the Scientific Software Collaborative component, we transform collaborative support efforts of today into routine self-support efforts of the future, freeing Extended Developer Support staff

to work on challenges associated with emerging technologies and new communities.

### 4.3 Scientific Software Collaborative

A majority of survey respondents expressed interest in receiving help with gateway building including choosing technologies and integrating new features and capabilities (e.g., visualization or computational tools, education support resources, etc.). The *Scientific Software Collaborative* is a crucial piece of the Institute's solution to address these needs. We have designed it around two guiding principles: Gateways should not be built as a series of one-off efforts but, at the same time, a single software solution will not fit all problems.

The goal of the *Scientific Software Collaborative* is to promote science gateways, simplify development and expand the capabilities of science gateways. By accessing components in an extensible software collaborative whose pieces can be used to build gateways, researchers will be able to focus their efforts on novel and challenging development specific to their research projects and user communities.

The *Scientific Software Collaborative* provides software solutions that facilitate the following:

*Discovery of gateways* for those looking to find existing gateways they might want to use or collaborate with.

*End-to-end solutions* to help developers from various scientific domains with minimal IT support who want an “out-of-the-box” hosted gateway for their application.

A “*use-what-you-need*” approach to help developers create their own customized gateways using selected components or Application Programming Interfaces (APIs).

*Software* for developers who want to actively extend open-source gateway software.

The high level framework that outlines the needs above can be seen in Figure 3. To facilitate the discovery of gateways the *Scientific Software Collaborative* works with *Community Engagement and Exchange* to promote existing science gateways. Currently, it can be difficult for interested parties to find operational science gateways that already meet their needs. We are creating a mechanism by which the community can publish information on active science gateways. While it will not be possible to maintain a comprehensive list of science gateways, an open registry made available by the Institute greatly improves the visibility of existing gateways. Similar to the project registry at [citizenscience.org](http://citizenscience.org) [41], we offer an easy-to-use, easy-to-manage, community-updated location for gateways to promote their projects and



help other projects connect with them. This initial project is one of the first offerings of this area and engage the community early on.

Simultaneously the *Scientific Software Collaborative* moves forward with the other components of the architecture. For gateway developers interested in an end-to-end solution, the *Scientific Software Collaborative* will include developer toolkits that use Docker [32,33]. Docker is an open platform to build and run distributed applications. A variety of executable images for developers to leverage as skeletons to rapidly deploy a secure and functional web presence for their gateway needs. Developers will also be able to create their own individualized containers, choosing just what they need. The Docker engine enables the *Scientific Software Collaborative* to distribute containers that comprise the application and its dependencies and run them as isolated processes in user space on a host operating system. Such “Dockerized” applications can run anywhere and be completely portable and reproducible.

The *Scientific Software Collaborative* will also provide individual components of systems and an API level of integration. APIs will be lightweight, event-driven, asynchronous, and easily consumed by every modern programming language. The APIs will be RESTful web services [34] that address the most common issues that face the gateway community today: information services, security services, execution services, data services, event services, and accounting services. We will base our API work on prior efforts defining the iPlant Agave API [35], the Apache Airavata API [36], and XSEDE APIs. This will provide an environment in which clients interact only with endpoints of components they need.

A hosting environment for Docker containers will be offered and, more generally, support for developers who need to develop, test, and create alpha releases of their gateways before migrating to their own platforms. This allows developers to create environments that suit their research needs and share them with collaborators. XSEDE-allocated projects will be able to use XSEDE hosting environments such as Quarry [37], Jetstream [38] and Comet [39].

Lastly, in order to continue to facilitate the creation of a community, the *Scientific Software Collaborative* enables software providers to promote their software tools, including APIs, to the wider science-gateway community. Software providers can contribute their own Docker containers and market their API capabilities. We encourage the community to provide feedback on developer toolkits and API endpoints and contribute improvements. The Institute will contribute expertise in interoperability by leveraging expertise in

the NSF EarthCube CINERGI project [36] and several NIH-funded information frameworks [37,38]. Because the *Scientific Software Collaborative* enables access to existing software packages, our efforts inherently support reuse. We also continually study community standards and, where appropriate, incorporate them from organizations such as the World Wide Web Consortium [43] and Open Grid Forum [44].

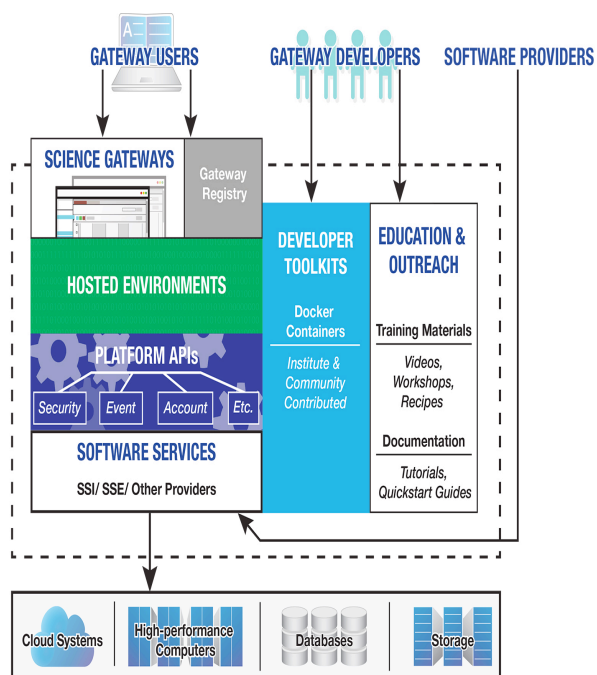


Figure 3: High-level design of the scientific software collaborative

The *Scientific Software Collaborative* will continue to evolve over time and engages the community early on to measure the impact and success on the community. The components of the *Scientific Software Collaborative* is leveraged by the *Incubator* and *Extended Developer Support* to help researchers build and improve their science gateways. We work closely with *Community Engagement and Exchange* to offer comprehensive documentation and outreach services specific to the framework and with *Workforce Development* to enable students to learn about the software components and improve their programming skills.

#### 4.4 Community Engagement and Exchange

The focus groups funded by our conceptualization grant and prior research [11,47] revealed that community members are eager to connect to and learn from their colleagues. The primary goal of *Community Engagement and Exchange*, therefore, is to facilitate

interaction within the community by providing multiple venues for learning, sharing, demonstrating, connecting, and community building. A second goal is to help community members learn about our variety of services. A final goal is to solicit input and feedback from the community to ensure that all Institute services remain relevant, useful, and effective.

Community Engagement and Exchange activities include a moderated discussion forum; a showcase of successful gateways that captures and shares lessons learned; postings of relevant gateway-related news and media coverage, events, academic publications, job openings, and websites; and a curated blog with guest authors on topics drawn from community expertise. Online communities were selected by nearly half (47%) of our survey's application-creator respondents as their favorite method for keeping up-to-date on relevant technologies, but we will also test newer approaches.

An annual conference is our primary face-to-face event for supporting professional development. Structured around the pressing needs and primary interests of the community, our conference includes presentations, workshops, exhibits, opportunities for networking, and an "Open Space" [48] to accommodate emerging issues. As part of this conference, we will publish proceedings and special journal issues with gateway groups throughout the world. During the year, the Institute also supports professional development through an online symposium series, workshops, and on-site training by request. Our survey indicated that these were the most popular forms of training for technology development staff.

An outreach program focuses on the formation of campus-based, gateway-developer groups (often in research IT units). Successful development of these campus-based groups can make the many research projects relying on cyberinfrastructure and related expertise more cost effective to execute on campuses. These groups also build stable career paths for developers.

In addition, the Institute wants to build connections among gateways and extend the reach of other independently funded, complementary NSF initiatives (e.g., XSEDE's Extended Support for Science Gateways, the Gateway Computing Environments workshops, existing gateway platforms, NSF SI<sup>2</sup> projects, Science and Technology Centers, Engineering Research Centers). We are actively soliciting organizations to become affiliates or partners so that we can help call attention to the resources that they offer while finding synergies with our own services.

We refine and adapt our services iteratively and follow a four-part strategy to obtain community

feedback: interviews, group-based input, online surveys, and social networks. In many cases, the collaboration with our external evaluator will simultaneously support Institute assessment. Social networking feedback tools allow customers to pose questions, discuss ideas, report problems, offer praise and suggestions, and prioritize.

#### 4.4 Workforce Development

Workforce gaps and shortages are problems looming throughout cyberinfrastructure and computational science and engineering efforts [49] and are major concerns for gateway developers. Workforce Development, therefore, aims to increase the pipeline of young developers and taps the unrealized potential of students from underrepresented groups.

*Education Using Gateways.* Fellowship and internship programs support talented students who commit to learning open-source, Web, and distributed-computing software skills that underlie science gateways. A second program connects with STEM professors who can integrate gateway science into their courses. This is particularly important at smaller colleges and universities, including Minority Serving Institutions (MSIs), that do not have the computational expertise or computing power on campus. As gateway groups grow on campuses, this will lead to job boards and opportunities for well-trained students. Students in computer science will become involved directly with gateway software development and implementation. Students in STEM fields will become better prepared to use computational methods in their work and when pursuing advanced degrees.

*Institute Interconnections.* The web is a natural platform for student involvement, and there is clear alignment between Workforce Development and the other Institute areas. Students and new developers will work with the Scientific Software Collaborative, getting involved with open-source science gateway and related software projects. Students can find problems and contribute solutions (such as bug fixes and improved documentation) without advanced training, while demonstrating their skills, investigating the latest technologies, and improving their online portfolios. Such work may include substantial extension and development of science gateways and enable students to conduct publishable scientific research earlier in their studies. Faculty working with Workforce Development to revise courses find assistance in curricula developed in coordination with Community Engagement and Exchange. Finally, Extended Developer Support works with new developers, creating internships with gateways that need support and with Institute mentors.



*Clients and Partners.* These ambitions are built on a concrete plan for action. The first, most important action will be to connect the Institute to clients and partners who help identify the students and faculty who serve, respectively, as young developers and gateway classroom users. We accomplish this through national searches and partnerships with the National Organization for the Professional Advancement of Black Chemists and Chemical Engineers (NOBCChE) and the Association of Computer/Information Sciences and Engineering Departments at Minority Institutions (ADMI). Partnerships with NOBCChE and ADMI were initiated during the Institute's conceptualization grant. Future efforts will extend to other appropriate societies in later project years.

*Implementation.* First, we will establish an Institute center for training and education on the campus of Elizabeth City State University (ECSU), a Historically Black University. The center will offer a vigorous schedule of on-site and virtual training conducted by the Institute and the larger gateway community. Second, we will develop training and course curricula about science gateway technologies and make them available as publicly accessible, online materials.

## 5. Future Work

The Institute is a service organization and dynamically reacts to experiences and feedback on all the diverse services offered. As one of the first NSF Software Institutes, we are offering some very novel services and expect that there will be many changes as we gain experience. We look forward to seeing the blossoming and return on investment of the Institute's many components – successful Incubator cohorts launching gateways and advising others, case studies from Extended Developer Support projects, searchable listings of hundreds of functional science gateways available for use through the Scientific Software Collaborative, a highly attended, dynamic annual conference and a workforce development effort that turns out great students with exciting career opportunities and science gateways as an important component of curricula from the diverse communities using services.

The five areas of the institute are working closely together and are leveraging each other in significant ways. For example, participants in workforce development internships will be paired with staff in Extended Developer Support to work on active science gateway projects. They may also assist with design and population of the Software Collaborative and gateway listing.

Finally, we establish strong collaborations internationally, with the International Coalition on

Science Gateways, the Virtual Research Environments interest group in the Research Data Alliance and continued co-publishing with the International Workshop on Science Gateways and the International Workshop on Science Gateways-Australia.

## Acknowledgements

We would like to acknowledge the assistance of Stephanie Sides for significant editorial contributions.

## 7. References

- [1] Cyberinfrastructure Framework for the 21st Century Science and Engineering Program of the National Science Foundation, [http://www.nsf.gov/about/budget/fy2012/pdf/40\\_fy2012.pdf](http://www.nsf.gov/about/budget/fy2012/pdf/40_fy2012.pdf).
- [2] <http://sciencegateways.org/resources/reports/>.
- [3] Cyberinfrastructure for Engineering Research and Education, <http://www.nsf.gov/eng/general/Workshop/cyberinfrastructure/index.jsp>.
- [4] Identifying Major Scientific Challenges in the Mathematical and Physical Sciences and Their Cyberinfrastructure Needs: A Workshop Funded by the National Science Foundation, April 21, 2004, [http://www.nsf.gov/attachments/100811/public/Cyberscience\\_Final4.pdf](http://www.nsf.gov/attachments/100811/public/Cyberscience_Final4.pdf).
- [5] Novotny, Mark A., David Ceperly, Chakram S. Jayanthi, and Richard M. Martin. Materials Research Cyberscience Enabled by Cyberinfrastructure, 2004, <http://www.nsf.gov/mps/dmr/csci.pdf>.
- [6] Berman, Francine, and Henry Brady. Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences, 2005, [http://ucdata.berkeley.edu/pubs/CyberInfrastructure\\_FINAL.pdf](http://ucdata.berkeley.edu/pubs/CyberInfrastructure_FINAL.pdf).
- [7] XSEDE: The Extreme Science and Engineering Discovery Environment, Quarterly Report for 2014, July 1, 2014 through September 30, 2014, <http://hdl.handle.net/2142/73169>.
- [8] Harmon, Amber. Saying Goodbye to the Command Line, ISGTW: International Science Grid This Week. October 2, 2013, <http://www.isgtw.org/feature/saying-goodbye-command-line>.
- [9] Wilkins-Diehr, N., Gesing, S., and Kiss, T. Special Issue on Science Gateway Workshops 2013 (SGW 2013). Concurrency Computation: Practice and Experience, 27(2):253–506, 2015.
- [10] Lengauer, C., Bouge, L., Silva, F., Li, Z., Li, K., Gesing, S. and Wilkins-Diehr, S. Special Issue: Combined Special Issues on EuroPar 2014, Recent advances in parallel computing and distributed network (U-Science 2014) and Science gateway workshops (SGW 2014). Concurrency Computation: Practice and Experience, 27(16):4073–4490, 2015.
- [11] Nancy Wilkins-Diehr. 2011. A history of the TeraGrid science gateway program: a personal view. In Proceedings of the 2011 ACM workshop on Gateway computing environments (GCE '11). ACM, New York, NY, USA, 1-12. DOI=10.1145/2110486.2110488

- [12] C. Catlett et al., "TeraGrid: Analysis of Organization, System Architecture, and Middleware Enabling New Types of Applications," *Advances in Parallel Computing*, Lucio Grandinetti, ed., IOS Press, ch. 3, 2008, pp. 225–249.
- [13] Wilkins-Diehr, N., J. Alameda, K. Droegemeier, and D. Gannon, D. 2008. The Importance of Long-term Science and Engineering Infrastructure for Digital Discovery. White paper. Retrieved 5/13/15 from <http://www.sciencegateways.org/sgwwhitepaper>.
- [14] Lawrence, Katherine A., and Nancy Wilkins-Diehr. Roadmaps, Not Blueprints: Paving the Way to Science Gateway Success. In *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the Campus and Beyond*, p. 40. ACM, 2012.
- [15] Lawrence, K. A., M. Zentner, N. Wilkins-Diehr, J. A. Wernert, M. Pierce, S. Marru, and S. Michael. 2015. Science Gateways Today and Tomorrow: Positive Perspectives of Nearly 5000 Members of the Research Community. *Concurrency and Computation: Practice and Experience* (in press).
- [16] Wenger, Etienne, Richard McDermott, and William M. Snyder. *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Harvard Business School Press, 2002.
- [17] Shih, Tse-Hua, and Xitao Fan. Comparing Response Rates in E-mail and Paper Surveys: A Meta-analysis." *Educational Research Review* 4, no. 1 (2009): 26-40.
- [18] Dykema, Jennifer, John Stevenson, Lisa Klein, Yujin Kim, and Brendan Day. Effects of E-mailed Versus Mailed Invitations and Incentives on Response Rates, Data Quality, and Costs in a Web Survey of University Faculty. *Social Science Computer Review* 31, no. 3 (2013): 359-370.
- [19] Niu, Xi, Bradley M. Hemminger, Cory Lown, Stephanie Adams, Cecelia Brown, Allison Level, Merinda McLure, Audrey Powers, Michele R. Tennant, and Tara Cataldo. National Study of Information Seeking Behavior of Academic Researchers in the United States. *Journal of the American Society for Information Science and Technology* 61, no. 5 (2010): 869-890.
- [20] Cohen, Susan. What Do Accelerators Do? Insights from Incubators and Angels. *Innovations* 8, no. 3-4 (2013): 19-25.
- [21] Apache: <http://www.apache.org/foundation/governance/>, <http://www.apache.org/foundation/governance/orgchart>.
- [22] Marru, Suresh, Lahiru Gunathilake, Chathura Herath, Patanachai Tangchaisin, Marlon Pierce, Chris Mattmann, and Raminder Singh. Apache Airavata: A Framework for Distributed Applications and Computational Workflows. In *Proceedings of the 2011 ACM Workshop on Gateway Computing Environments*, pp. 21-28. ACM, 2011.
- [23] McLennan, Michael, and Rick Kennell. HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering. *Computing in Science & Engineering* 12, no. 2 (2010): 48-53.
- [24] <https://crc.nd.edu/index.php/research/gateways>.
- [25] Osterwalder, Alexander, and Yves Pigneur. *Business Model Generation: A Handbook For Visionaries, Game Changers, and Challengers*. Wiley, 2010.
- [26] Maron, Nancy, K. Kirby Smith, and Matthew Loy. *Sustaining Digital Resources: An On-the-ground View of Projects Today: ITHAKA Case Studies in Sustainability*. Ithaka S+ R, 2009.
- [27] Madhavan, Krishna, Michael Zentner, and Gerhard Klimeck. *Learning and Research in the Cloud*. *Nature Nanotechnology* 8, no. 11 (2013): 786-789.
- [28] Michael Zentner (2013), On Safari in nanoHUB: The Hunt for Prototypical Users, <https://hubzero.org/resources/990>.
- [29] Galligan, Finbar, and Sharon Dyas-Correia. Altmetrics: Rethinking the Way We Measure. *Serials Review* 39, no. 1 (2013): 56-61.
- [30] SCI-BUS (SCientific gateway Based User Support), FP7 Project No. RI-283481, <http://www.sci-bus.eu>.
- [31] Kacsuk, Peter, Zoltan Farkas, Miklos Kozlovsky, Gabor Hermann, Akos Balasko, Krisztian Karoczkai, and Istvan Marton. WS-PGRADE/gUSE Generic DCI Gateway Framework for a Large Variety of User Communities. *Journal of Grid Computing* 10, no. 4 (2012): 601-630.
- [32] Merkel, Dirk. Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux Journal* 2014, no. 239 (2014): 2.
- [33] <https://hub.docker.com>.
- [34] Richardson, Leonard, and Sam Ruby. *RESTful Web Services*. O'Reilly Media, Inc., 2008.
- [35] Dooley, Rion, and Matthew R. Hanlon. *Recipes 2.0: Building for Today and Tomorrow*. *Concurrency and Computation: Practice and Experience* (2014).
- [36] Pierce, Marlon, Suresh Marru, Borries Demeler, Raminderjeet Singh, and Gary Gorbet. The Apache Airavata Application Programming Interface: Overview and Evaluation with the UltraScan Science Gateway. In *Proceedings of the 9th Gateway Computing Environments Workshop*, pp. 25-29. IEEE Press, 2014.
- [37] <https://portal.xsede.org/iu-quarry>.
- [38] <http://www.jetstream-cloud.org/>.
- [39] <http://comet.sdsc.edu>.
- [40] <http://earthcube.org/group/cinergi>
- [41] <http://www.neuinfo.org/>.
- [42] <https://www.nitrc.org/>.
- [43] <http://www.w3.org>
- [44] <https://www.ogf.org>.
- [45] Bonney, Rick, Caren B. Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V. Rosenberg, and Jennifer Shirk. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience* 59, no. 11 (2009): 977-984.
- [46] <https://tis.xsede.org/>.
- [47] Wilkins-Diehr, Nancy, and Katherine A. Lawrence. Opening Science Gateways to Future Success: The Challenges of Gateway Sustainability. In *Gateway Computing Environments Workshop, 2010*, IEEE, pp. 1-10.
- [48] Owen, Harrison. *Open Space Technology: A User's Guide*. Berrett-Koehler Publishers, 2008.
- [49] Advanced Scientific Computing Advisory Committee Subcommittee Letter, Assessment of Workforce Development Needs in Office of Science Research Disciplines, [http://science.energy.gov/~media/ascr/ascac/pdf/charges/AS\\_CAC\\_Workforce\\_Letter\\_Report.pdf](http://science.energy.gov/~media/ascr/ascac/pdf/charges/AS_CAC_Workforce_Letter_Report.pdf)